5Л

УДК 004.89

В.В. Литвин, Р.Р. Даревич, Д.Г. Досин, Н.В. Шкутяк
Національний університет «Львівська політехніка», м. Львів, Україна
Фізико-механічний інститут НАН України ім. Г.В. Карпенка, м. Львів, Україна
vasyll@ukr.net, darevych@ipm.lviv.ua, dosyn@ipm.lviv.ua

Проектування інтелектуальних агентів прийняття рішень в просторі ознак із використанням онтологічного підходу

У статті розглядається проблема побудови метрики для знаходження релевантних прецедентів інтелектуальним агентом, який функціонує в просторі ознак. Пропонується зважувати ознаки коефіцієнтами їх важливості, які зберігаються в онтології предметної області. Розглянуто способи визначення цих коефіцієнтів.

Вступ

Сучасний рівень розвитку інтелектуальних систем прийняття рішень відбувається у двох напрямках розроблення інтелектуальних агентів (ІА) [1]:

– інтелектуальні агенти планування своєї діяльності;

– інтелектуальні агенти, засновані на прецедентах.

Перший напрямок використовує відомий підхід прийняття рішень як де дуктивний процес з використанням логічного виведення, заснованого на правилах. Задача, яку розв’язує інтелектуальний агент, полягає в досягненні стану мети. Часто інтелектуальність агента полягає в тому, щоб цей стан досягнути з якомога найменшими затратами своїх ресурсів. Розв’язок такої задачі зводиться до пошуку шляху у просторі станів. Для оцінки релевантності станів, що лежать на шляху, використовуються метрики, які задаються у вигляді евристичних функцій. Такі евристичні функції, а отже як наслідок і метрика, можуть базуватися на онтології предметної області (ПО), в межах якої функціонує ІА.

З часом було усвідомлено, що дедуктивна модель моделює один з найбільш рідкісних підходів, який використовує ІА під час вирішення проблеми. Ідея виведення за правилами є привабливою, тому що вона ґрунтується на добре формалізованих поняттях, для яких існують наукові методи, що довели свою застосовність і які дозволяють отримати рішення, що не вимагають доведень. Однак навколишній світ складний. Існує багато слабо формалізованих завдань, для яких, можливо, будуть знайдені рішення. Крім того, існує ряд завдань, для яких ніколи не буде знайдено формальне рішення (судочинство, медицина). Актуальність проблеми обумовлена і чисельністю таких завдань, і практичною потребою знайти хоча б одне яке-небудь відповідне рішення там, де через відсутність строгого формалізованого методу не можна знайти всі або найомтимальніші рішення.

Насправді, замість того, щоб вирішувати кожну задачу, виходячи з первинних принципів, ІА часто аналізує ситуацію в цілому і згадує, які рішення ухвалювалися раніше в подібних ситуаціях. Потім він або безпосередньо використовує ці рішення, або, за необхідністю, адаптує їх до обставин, що змінилися для конкретної проблеми. Моделювання такого підходу до вирішення проблем, заснованого на досвіді минулих

ситуацій, привело до появи технології логічного виведення, заснованого на прецедентах (англійською – Case-Based Reasoning, або CBR), і надалі – до створення програмних продуктів, що реалізовують цю технологію.

У ряді ситуацій метод виведення за прецедентами має серйозні переваги в порівнянні з виведенням, що ґрунтується на правилах, і особливо ефективний, коли:

– основним джерелом знань про завдання є досвід, а не теорія;

– рішення не унікальні для конкретної ситуації і можуть бути використані в інших випадках;

– метою є не гарантоване вірне рішення, а краще з можливих.

Таким чином, виведення, засноване на прецедентах, є методом побудови інтелектуальних систем, які приймають рішення щодо даної проблеми або ситуації за наслідками пошуку аналогій, що зберігаються в базі прецедентів.

Як бачимо, для обох класів інтелектуальних систем прийняття рішень необхідна метрика. У першому випадку – для оцінювання релевантності станів, у другому випадку – для оцінювання релевантності прецедентів. Від способу визначення цієї метрики напряму залежить ефективність роботи ІА. На наш погляд, такий спосіб повинен базуватись на онтологіях ПО, в межах яких функціонує ІА.

Метою цієї роботи є побудова метрики на основі онтологій ПО для ефективного функціонування ІА.

Класифікація задач

Проаналізувавши клас задач, які розв’язуються інтелектуальною системою, можна зробити висновок, що всі задачі можна поділити на два величезні підкласи, виходячи із самої постановки задачі. Так, існує клас задач, для яких суттєве значення понять (властивостей). Сюди відносяться задачі діагностики захворювань, кваліфікації явищ на основі збору даних тощо. Такі задачі назвемо ознаковими. Для іншого класу задач не є суттєвим значення понять, а скоріше їх семантика або частотність. Зокрема, сюди можна віднести кластеризацію інформаційних ресурсів, класифікацію текстів згідно з УДК, пошукові системи. Такий клас задач назвемо семантичними задачами.

В результаті отримаємо поділ інтелектуальних систем за двома вимірами, так як це зображено на рис. 1.

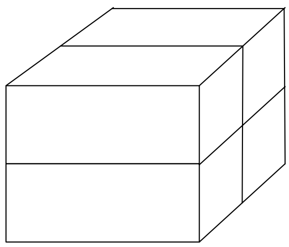


Рисунок 1 – Класифікація інтелектуальних систем прийняття рішень

Для ефективного функціонування ІА необхідно побудувати метрику, на основі якої визначати релевантність станів чи прецедентів. На наш погляд, побудова такої метрики напряму залежить від класу задач: семантичні вони чи ознакові. Онтологія першого класу більш спрямована на врахування семантики ПО. Онтологія другого класу – віднесення об’єктів до класів. Формально онтологію задамо як п’ятірку [2]

$$O = \langle C, R, F, W, L \rangle ,$$

де W – важливість понять C, L – важливість відношень R, F – інтерпретація цих понять та відношень.

Семантичні задачі (використовуються нагромаджувальні методи задання ваг термінів та відношень). До класу цих задач належать задачі, для яких є несуттєвим значення ознак об'єктів, що належать до класів онтології. Суть цих задач полягає у визначенні семантики інформації:

а) класифікація за УДК;
б) інформаційно-пошукові системи за контекстом;
в) кластеризація інформації (структуризація масиву інформації);
г) автоматичне реферування;
д) переклад з мови на мову.

Для семантичних задач метрика визначається як відстань в концептуальному графі між n-ою кількістю центрів ваг. Детально цей спосіб нами розглянуто у роботах [3], [4].

Тому у цій роботі детальніше розглянемо проблему побудови ефективних метрик для ознакових задач, а саме для пошуку релевантних прецедентів. Очевидно, що визначена для таких задач метрика успішно може використовуватися у інтелектуальних системах планування діяльності.

Ознакові задачі (використовуються ймовірні методи, експертні методи та методи на основі інтелектуального аналізу даних). До класу цих задач належать задачі для яких є суттєвим значення ознак об'єктів, що належать до класів онтології:

а) задачі медицини (діагностика захворювань);
б) прогнозування явищ;
в) прийняття рішень в активних середовищах.

Побудова метрики для ознакових задач

Нехай множина прецедентів $\boldsymbol{Pr} = \{\boldsymbol{Pr}_1, \boldsymbol{Pr}_2, \ldots, \boldsymbol{Pr}_N\}$ описується характеристиками (властивостями) $\boldsymbol{X} = \{x_1, x_2, \ldots, x_M\}$. D_i – домен властивості x_i. Для відображення властивостей ми використовуємо нове позначення $\boldsymbol{X}$, а не C, оскільки, як правило, ці значення є властивостями об'єктів, які належать класу C. Хоча можливі випадки, що окремі властивості тотожні класу. Це залежить від того, як був проведений онтологічний інжиніринг.

Релевантним $\boldsymbol{Pr}_k$ є прецедент, для якого відстань до поточної ситуації є найменшою, тобто

$$\boldsymbol{Pr}_k = \arg\min_i d_i$$

Тоді прецеденти розрізняються між собою в залежності від значень, які набувають ці властивості. Очевидно, що для розпізнавання прецеденту не потрібно значення всіх властивостей множини $\boldsymbol{X}$, а лише деякої її підмножини. Дійсно, якщо розглянути деяке дерево рішень, то окрема гілка від кореня дерева до листка описує окремий прецедент, а значення властивостей, які знаходяться на цій гілці достатньо для опису цього прецеденту. Тобто

$$\mathrm{Pr}_i \leftrightarrow X_i = \left\{ x_{i_1} = z_{i_1}, x_{i_2} = z_{i_2}, \ldots, x_{i_k} = z_{i_k} \right\}, \text{ де } z_{i_j} \in D_{i_j}.$$

З іншого боку, властивість, яка знаходиться на вищому рівні дерева рішень, повинна мати більшу вагу в адаптивній онтології. Оскільки для різних прецедентів властивості можуть знаходитися на різних рівнях дерева рішень, то коефіцієнти важливості понять насправді представляють собою вектори виміром кількості прецедентів, тобто $W = (W_1, W_2, \ldots, W_N)$, де W_i коефіцієнти важливості понять онтології, яка описує прецедент $\boldsymbol{Pr}_i$.

Тоді відстань між прецедентом $\boldsymbol{Pr}_i$ та поточною ситуацією S:

$$d_i = \sum_{l=1}^{k} w_{i_l} \cdot \varphi\left(z_{i_l}, z_{i_l}^S \right), \qquad (1)$$

де z_{i_l} значення властивості x_{i_l} прецедента ***Pr**ᵢ*, $z_{i_j}^S$ значення властивості x_{i_l} поточної ситуації *S*, w_{i_l} коефіцієнт важливості властивості x_{i_l} прецедента $\boldsymbol{Pr_i}$.

Розглянемо функцію $\varphi(x,y)$. Очевидно, що *x* – може бути діапазоном, тобто нечіткою множиною, де *D* – універсальна множина $x \subseteq D$; числовим значенням або нечисловим значенням. В залежності від цього $\varphi(x,y)$ визначається по-своєму:

$$\varphi(x,y)=\begin{cases} 1-\mu_x(y),\ x-\text{нечітка множина},\\ \lambda\cdot|x-y|,\ x,y-\text{числові значення},\\ 1-\mu(x,y),\ x,y-\text{нечислові значення}. \end{cases} \qquad (2)$$

де $\mu_x(y)$ – коефіцієнт впевненості того, що y належить нечіткій множині x, λ – числова величина, яка залежить від ПО, щоб $\lambda\cdot|x-y|\in[0,1]$, $\mu(x,y)\in[0,1]$ – нечітка величина подібності значень x та y. Наприклад $\mu(x,y)=1$, якщо $x=y$, $\mu(x,y)=0{,}9$, якщо $x\approx y$, $\mu(x,y)=0$, якщо $x\neq y$.

Розглянемо методи задання початкових коефіцієнтів важливості понять та зв’язків та їх модифікацію в процесі функціонування інтелектуального агента, який використовує цю онтологію.

Методи задання початкових ваг (коефіцієнтів важливості) понять та зв’язків:

1. За рахунок експертних оцінок.
2. Присвоєння випадковим чином.
3. За рахунок аналізу (статистичного, інтелектуального) інформаційних джерел, які описують ПО, в якій функціонує ІА.

Окрім того, ці ваги можуть мати обмеження на величину, наприклад їх значення знаходяться у проміжку [0,1] (ймовірнісні методи) або без обмежень на величину (нагромаджувальні методи).

Розглянемо визначення ваг термінів онтології на основі інтелектуального аналізу даних, а саме на основі побудови дерева рішень (ДР). Як відомо, ознакові задачі дозволяють для пошуку релевантних прецедентів будувати ДР. Однак ДР не є панацеєю, оскільки згадувані ознаки, що лежать на відповідній гілці, що задає прецедент, не гарантують врахування повної множини ознак, які необхідно врахувати для знаходьження релевантного прецедента. Тому комбінований метод (побудова ДР – розвинення ДР на основі онтології) виглядає перспективним. Нами пропонується використовувати ДР для визначення ваг базових термінів, які задають деякий прецедент, а потім на основі онтології ПО розвинути отримані ваги на всю онтологію для відповідного прецедента. Тоді для пошуку релевантного прецедента використовувати значення тих *n* концептів, які для відповідного прецедента мають найбільші ваги.

Відомо, що отримати правила пошуку релевантних прецедентів можна на основі побудови дерева рішень.

Розглянемо гілку дерева. Вершини (ознаки) цієї гілки знаходяться на *k* рівнях. Очевидно, що чим вищий рівень, тим значуща ознака, яка на цьому рівні знаходиться. Ця евристична думка має бути відображена в значеннях ваг цих ознак. Крім того, пропонується ці ваги пронормувати, тобто щоб їх сума для кожного прецедента (гілки) була рівна 1.

Розглянемо два способи визначення ваг базових ознак, які задовольняють вищеописані два припущення.

1-й спосіб. Арифметичні ваги. Визначаються як відношення різниці (*k*+1) рівня

дерева та рівня, на якому знаходиться ознака, до суми всіх рівнів гілки, тобто базуються на сумі арифметичної прогресії:

$$w_i = \frac{k+1-i}{\sum_{j=1}^{k} j} = \frac{k+1-i}{\frac{(1+k)k}{2}}.$$

2-й спосіб. Геометричні ваги. Базуються на сумі геометричної прогресії:

$$w_i = \frac{2^{k-i}}{2^k - 1}.$$

Так, для гілки рівня $k = 4$ отримаємо:

Рівень знаходження ознаки в ДР	Арифметична вага	Геометрична вага
1-й	$w_1 = \frac{2}{5}$	$w_1 = \frac{8}{15}$
2-й	$w_2 = \frac{3}{10}$	$w_2 = \frac{4}{15}$
3-й	$w_3 = \frac{1}{5}$	$w_3 = \frac{2}{15}$
4-й	$w_4 = \frac{1}{10}$	$w_4 = \frac{1}{15}$

Легко бачити, що $\sum_{i=1}^{4} w_i = 1$ для обох способів визначення ваг.

Отримані на основі ДР ваги назвемо вагами базових ознак прецедента і позначимо таку множину ваг W_B. Розвинемо їх на всю онтологію ПО, використовуючи таксономію понять онтології та інтерпретацію термінів, тобто:

$$W_B \xrightarrow{O} W.$$

Зауважимо, що наразі ми розглядаємо лише один конкретний прецедент. Насправді вага терміна є вектор розмірності кількості прецедентів, про що зазначалось вище.

Розмноження ваг на всю онтологію залежить від визначення (аксіоматизації) класів, їх ієрархії (вертикальний зв'язок) та горизонтальних зв'язків.

Тепер впорядкуємо множину ознак відповідно до їх ваг, тобто $w_1 \geq w_2 \geq w_3 \geq \ldots$. Для визначення відстані між прецедентом та поточною ситуацією за допомогою формули (1) візьмемо перші N ознак.

Приклад функціонування інтелектуального агента

Розглянемо приклад функціонування ІА в області медицини, а саме захворювань опорно-рухового апарата. За основу візьмемо результати досліджень, що одержали у Львівському Національному медичному університеті ім. Данила Галицького Е.Х. Заремба й О.О. Зімба. З деякими з них можна ознайомитися в [5]. Прецедентами є можливі захворювання. Для прикладу розглянемо три хвороби: ***Pr*** = {Ревматизм, Артрит, Подагра}.

Провівши онтологічний інжиніринг ПО, ми одержали множину властивостей, які необхібно дослідити та їх важливість в залежності від прецеденту. Наведемо деякі з них: ***X*** = {*Ступінь недиференційованої дисплазії сполучної тканини, Концентрація*

ендотеліну-1 у плазмі крові, Температура, Біль у суглобі}. Розглянемо детальніше ці ознаки. Домен першої є нечислові значення, другої і третьої – відрізки, четвертої – бінарне значення. Так $D_{С_н_д_с_т}$ = {слабке, середнє, сильне}, $D_{К_е_п_к}$ = [0;3,5], $D_{Температура}$ = [36,42], $D_{Біль_у_суглобі}$ = {Так(1),Немає(0)}. Значення цих ознак для ревматизму: *Температура*$_{\text{Ревматизм}}$ = = {(36|0), (37|0,6), (38|1), (39|1), (40|0,7)} – нечітка множина, *Біль у суглобі*$_{\text{Ревматизм}}$ = «Так(1)».

Важливість цих властивостей (коефіцієнти *W*) ми одержали методами статистичного аналізу (було досліджено 110 пацієнтів). Так для захворювання на ревматизм важливими симптомами є: наявність лади (w = 0,7), астенічна статура (w = 0,65), тонка шкіра (w = 0,45), сколіоз (w = 0,35). Перший симптом являє собою бінарну величину, що приймає значення з множини {Так(1),Немає(0)}, дві наступні – нечіткі величини, що приймають значення з відрізка [0;1] (зрозуміло, що 0 – відсутність симптому, 1 – повна впевненість у його наявності) і четвертий симптом теж нечітка змінна, тільки її носій – відрізок [0;3]. Для ревматизму приймаємо такі значення симптомів: $z_{лади}$ = 1, $z_{ас}$ = 1, $z_{тк}$ = 1, $z_{сколіоз} \in$ [1, 5; 3].

Нехай при обстеженні пацієнта *Y* ми одержали такі значення властивостей $y_{лади}$, $y_{ас}$, $y_{тк}$, $y_{сколіоз}$. Тоді відстань до прецедента, що цей пацієнт хворий ревматизмом, виміряється як:

$$d_{\text{ревматизм}} = 0{,}7\cdot|1-y_{лади}| + 0{,}65\cdot|1-y_{ас}| + 0{,}45\cdot|1-y_{тк}| + 0{,}35\cdot|1-\mu_{[1,5;3]}(y_{сколіоз})|$$

Аналогічним чином можна виміряти відстань до інших захворювань (артрит, подагра). Пацієнт хворий тим захворюванням, відстань до якого є найменшою.

Висновки

У роботі розглянуто модель функціонування ІА для задач, для яких важливі значення ознак. Побудована метрика для пошуку релевантних прецедентів, яка базується на онтології ПО. Це надало можливість підвищити ефективність функціонування таких ІА. Розглянуто способи визначення ваг понять онтології.

Література

1. Інтелектуальні системи, базовані на онтологіях / [Д.Г. Досин, В.В. Литвин, Ю.В. Нікольський, В.В. Пасічник]. – Львів: «Цивілізація», 2009. – 414 с.
2. Литвин В.В. Мультиагентні системи підтримки прийняття рішень, що базуються на прецедентах та використовують адаптивні онтології / В.В. Литвин // Радіоелектроніка, Інформатика, Управління. – Запоріжжя, 2009 . – № 2(21). – С. 120–126.
3. Даревич Р.Р. Метод автоматичного визначення інформаційної ваги понять в онтології бази знань / Р.Р. Даревич, Д.Г. Досин, В.В.Литвин // Відбір та обробка інформації. – 2005. – Вип. 22(98). – С. 105-111.
4. Оцінка подібності текстових документів на основі визначення інформаційної ваги елементів бази знань / Р.Р. Даревич, Д.Г. Досин, В.В. Литвин, З.Т. Назарчук // Искусственный интеллект. – 2006. – № 3. – С. 500-509.
5. Заремба Є.Х. Ендотеліальна дисфункція у хворих на ревматизм із недиференційованою дисплазією сполучної тканини / Є.Х. Заремба, О.О. Зімба // Український ревматологічний журнал. – 2009. – № 4(38). – С. 32-34.

V.V. Lytvyn, R.R. Darevych, D.G. Dosyn, N.V. Shkutiak

Planning of Intellectual Decision-making Agents in an Attributes Space with the Use of Ontological Approach

In the paper the problem of metrics construction for searching of relevant precedents by an intellectual agent which operate in the attributes space is considered. The attributes is proposed to weigh by the coefficients of their importance, which is storage in domain ontology. The approaches of determination of these coefficients are considered.

Стаття надійшла до редакції 01.06.2010.